A MARKOV CHAIN MODEL FOR THE PROBABILITY OF PRECIPITATION OCCURRENCE IN INTERVALS OF VARIOUS LENGTH

JAMES E. CASKEY, JR.

U.S. Weather Bureau, Washington, D.C.

[Manuscript received August 29, 1962; revised March 21, 1963]

ABSTRACT

Theoretical probabilities derived from a simple Markov chain model are found to agree closely with the empirical values of the probability of precipitation occurrence in intervals of various length at Denver, Colo.

1. INTRODUCTION

The purpose of this note is to point out that the probability distributions shown in tables 1 and 2 and figures 1 and 2 of the preceding paper by Topil [1] are closely approximated by a simple Markov chain model. Gabriel and Neumann [2] showed that the Markov chain model holds well for daily rainfall occurrences at Tel Aviv. They derived from the model the probability formulas for various features of the rainfall distribution, but did not give the explicit formula for the probability of rainfall occurrence as a function of the length of period, the distribution considered by Topil. Although the formula can be derived from their equation (12), a simple direct derivation is given in section 2. Theoretical probability curves computed from the formula are compared with Topil's empirical results for Denver, Colo., in section 3.

2. DERIVATION OF MODEL FORMULA

The immediate objective is to express the probability P_n of precipitation occurrence in an interval of n days in terms of the probability P_{n-1} of precipitation occurrence in $n-1$ days. If ρ_{n-1} is the conditional probability of a day being wet following a period of $n-1$ dry days (whose probability is $1-P_{n-1}$), then the probability $1-P_n$ of n dry days is given by

$$1-P_n=(1-P_{n-1})(1-\rho_{n-1}) \tag{1}$$

Expansion of (1) and rearrangement of terms give a recursive formula for P_n:

$$P_n=(1-\rho_{n-1})P_{n-1}+\rho_{n-1} \tag{2}$$

If ρ_{n-1} varies with n, then its value for each n over the range of interest must be estimated from the data sample. In this case, P_n may as well be estimated directly from the data sample, as done by Topil, rather than from (2). It is of interest, therefore, to let ρ_{n-1} be determined as a parameter given by the assumption of simple models of precipitation occurrence. For example, the choice of a random model, in which the probability of a day being wet is independent of what occurred on any preceding day, gives $\rho_{n-1}=P_1$ for all $n>1$. Because of the known persistence in precipitation data, however, this choice hardly seems plausible, and indeed the use of equation (2) with $\rho_{n-1}=P_1$ yields values of P_n that have a large positive bias relative to Topil's values for $n>1$.

The simple Markov chain, in view of its previous success (cf. Gabriel and Neumann [2]), seems to be a more promising choice since it takes some account of persistence. In the Markov chain model, the probability of precipitation on any day depends on whether or not precipitation fell the preceding day. Because the probability is independent of earlier days, $\rho_{n-1}=\rho_1$ for all $n>1$, where ρ_1 is the conditional probability that a day will be wet if the previous day was dry. Substitution of ρ_1 for ρ_{n-1} in equation (2) gives the basic recursive formula for the Markov chain model:

$$P_n=(1-\rho_1)P_{n-1}+\rho_1 \tag{3}$$

Successive substitution of the equations given by (3) for the series of integers 2, 3, . . ., $n-1$ yields an alternative formula:

$$P_n=1-(1-P_1)(1-\rho_1)^{n-1} \tag{4}$$

If P_n is to be computed for a series of successive integer values of n, equation (3) is convenient to use, but for a few isolated values of n, equation (4) is more convenient. In either case, values of ρ_1 and P_1 must be known in order to make the computations for P_n, $n\neq 1$. Both ρ_1 and P_1 may be estimated from the observed frequencies. Alternatively, as will be done in the next section, ρ_1 and τ_1 (where τ_1 is the conditional probability that a day will be wet if the previous day was wet) may be estimated from the observed frequencies; then P_1, the probability of any day being wet, can be computed from the identity

$$P_1=\rho_1(1-\tau_1+\rho_1)^{-1} \tag{5}$$

Note that in (5) if $\rho_1=\tau_1$ (i.e., previous day makes no difference), then $P_1=\rho_1$; this gives the random model which will be used later for comparison with the Markov chain model.

3. COMPARISON OF MODEL AND EMPIRICAL RESULTS FOR DENVER

The observed frequencies needed to compute ρ_1 and τ_1 for the four seasons at Denver, Colo., were kindly furnished by Topil [3] and are reproduced in tables 1 and 2. Table 1 gives the frequencies and the corresponding values ρ_1 and τ_1 when a wet day is defined as a day with a trace or more of precipitation. Table 2 gives similar information when a wet day is defined as a day with 0.01 inch or more of precipitation.

For the pair of values ρ_1 and τ_1 for each season [1] and for each definition of a wet day, P_1 was computed from equation (5). With P_1 and ρ_1 known, equation (4) gives a theoretical curve for the Markov chain values of P_n. In figures 1 and 2 such a curve [2] was drawn for each season and definition of wet day, and the corresponding observed data from Topil's [1] tables 1 and 2 (or if not tabulated, from the empirical curves in his figs. 1 and 2) were plotted for $n=1, 2, \ldots, 15$. The Markov chain curves fit the observed data well, as judged visually.

A comparison with Topil's observed values for periods of length 1 hr. to 12 hr. was not made in figures 1 and 2 because the assumption regarding persistence from one discrete day to the next, on which the Markov chain model is based, is meaningless for periods shorter than one day.

In figure 3, Topil's observed probabilities for $n=2, \ldots, 15$ for the four seasons and both definitions of wet day were plotted against the corresponding Markov chain values and also against values computed for random daily precipitation occurrences. The random model, which is obtained from equation (2) by putting $\rho_{n-1}=\rho_1=P_1$, would be appropriate if persistence of daily rainfall occurrence were negligible. Figure 3 shows the marked extent to which the Markov chain model improved the correlation between computed and observed probabilities for Denver over that given by the random model.

[1] Because of the marked changes in τ_1 and ρ_1 from month to month in some seasons, it would be desirable to make comparisons with Topil's results for each month. This was not done because the reliability of the monthly estimates of P_n for large n suffers from the small number of cases available in only 10 years of data. Accordingly the comparisons that follow are limited to four seasons. A logical way to select "precipitation seasons" would be to group months that have comparable values of P_1 and of ρ_1, the parameters of equation (4). However, since the present purpose is to make comparisons with Topil's results, his grouping of the months into the conventional winter, spring, summer, and autumn seasons was followed.

[2] The exponential form of equation (4) for the theoretical Markov chain probabilities could be very conveniently graphed on semilog paper, where it would be a straight line, for comparison with Topil's empirical values. This was not done because such a selection of coordinates unfortunately gives maximum resolution for the least reliable probability estimates (P_n, n large) and minimum resolution for the most reliable probability estimates (P_n, n small). For example, such a graph on 2-cycle semilog paper provides the same linear interval for the range $0.90\leq P_n\leq 0.99$ as it does for the much larger and more reliable range $0\leq P_n<0.90$.

TABLE 1.—*Number of wet days (trace or more) and all days classified by precipitation occurrence (trace or more) on preceding day, by seasons; and estimates of conditional probabilities of wet-day occurrences (for Denver, Colo., 1949–1958)*

	Preceding day	Actual day		Estimate of probability	
		Wet	Total	ρ_1	τ_1
Winter	Wet	127	250		0.508
	Dry	130	652	0.199	
Spring	Wet	284	445		0.638
	Dry	162	475	0.341	
Summer	Wet	330	513		0.643
	Dry	183	407	0.450	
Autumn	Wet	131	263		0.498
	Dry	129	518	0.199	

TABLE 2.—*Number of wet days (0.01 inch or more) and all days classified by precipitation occurrence (0.01 inch or more) on preceding day, by seasons; and estimates of conditional probabilities of wet-day occurrence (for Denver, Colo., 1949–1958)*

	Preceding day	Actual day		Estimate of probability	
		Wet	Total	ρ_1	τ_1
Winter	Wet	63	146		0.432
	Dry	91	756	0.120	
Spring	Wet	138	281		0.491
	Dry	138	639	0.216	
Summer	Wet	101	258		0.391
	Dry	159	662	0.240	
Autumn	Wet	68	154		0.442
	Dry	85	756	0.112	

4. CONCLUDING REMARKS

As has been emphasized by Gabriel and Neumann [2], the success of the Markov chain model in producing theoretical probabilities that agree closely with observed probabilities should not be interpreted as meaning that these distributions have been "explained." Knowledge that the Markov chain model fits certain observations well, merely gives confidence in its convenient use to compute many other properties of the observations, for example the several properties whose probability formulas are given by Gabriel and Neumann [2].

In view of its convenience for computations and its success in fitting daily data for Tel Aviv [2] and for Denver, the Markov chain model seems to deserve further investigation for a variety of synoptic-climatic regimes. How well, for example, would it fit the rainfall probability distributions for San Francisco, where Jorgenson [4] has shown persistence to be an important factor for 5 or 6 days? What is the effect of the assumption that ρ_1 and τ_1 are constants, especially throughout a season, when in fact they may undergo rather large month-to-month changes as shown for example by Eichmeier and Baten [5] in Michigan data? Studies beyond the scope of this brief

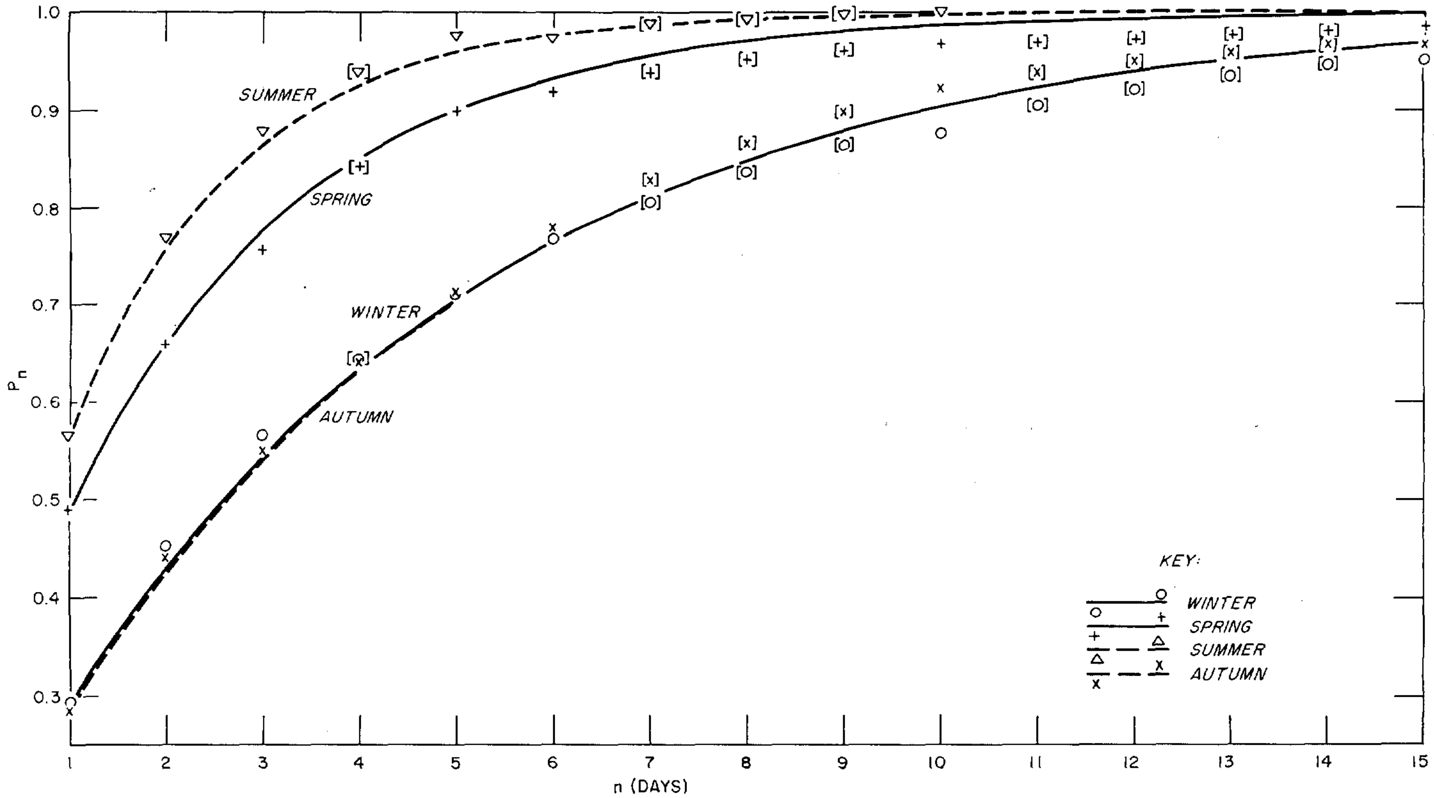

FIGURE 1.—Probability (P_n) of precipitation occurrences (trace or more) in periods of various length (n) at Denver, Colo. The curves represent the theoretical Markov chain model probabilities and the plotted values are Topil's [1] observed relative frequencies (bracketed values were obtained from his smoothed empirical curves).

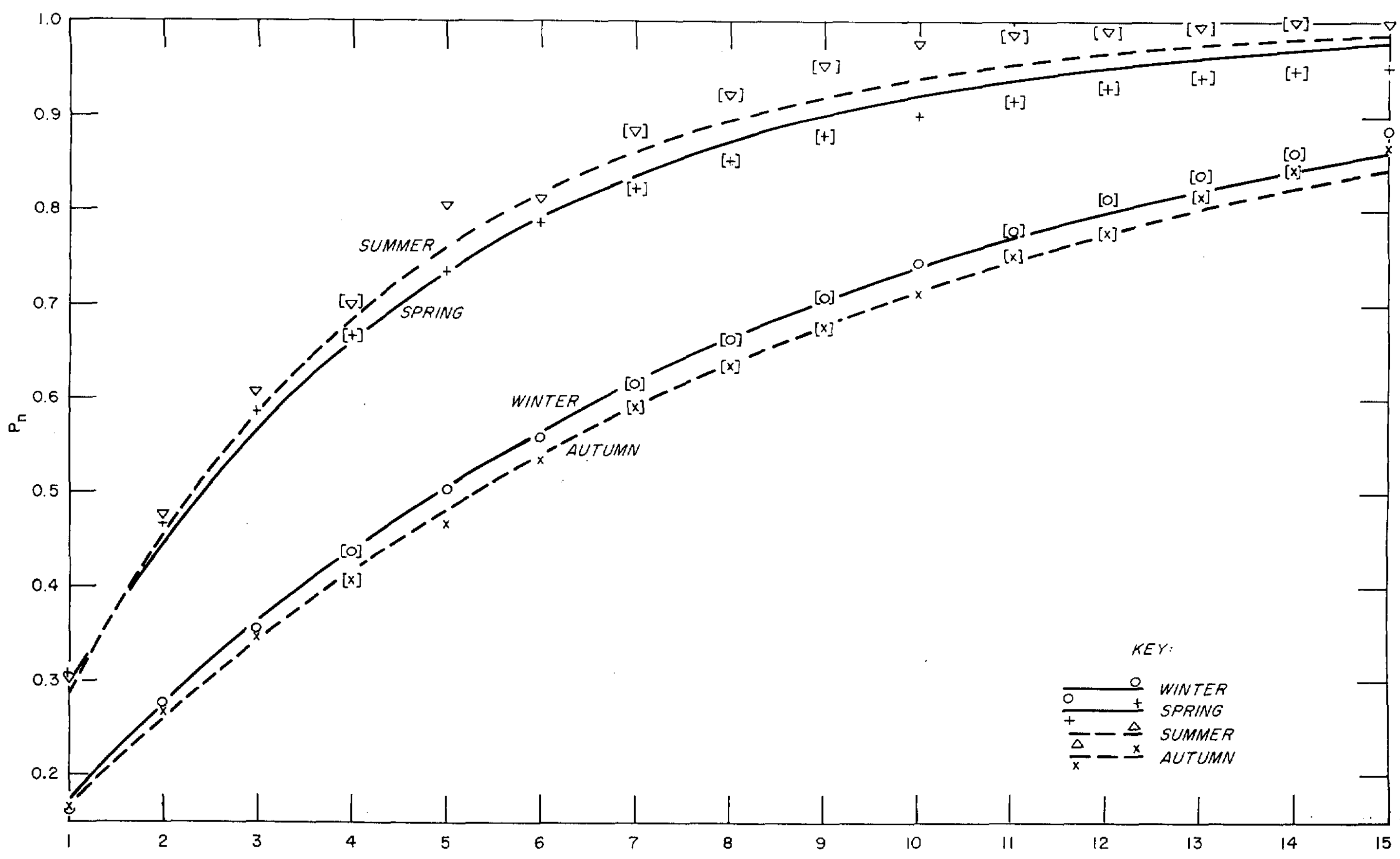

FIGURE 2.—Same as figure 1 except precipitation occurrence is defined as 0.01 inch or more.

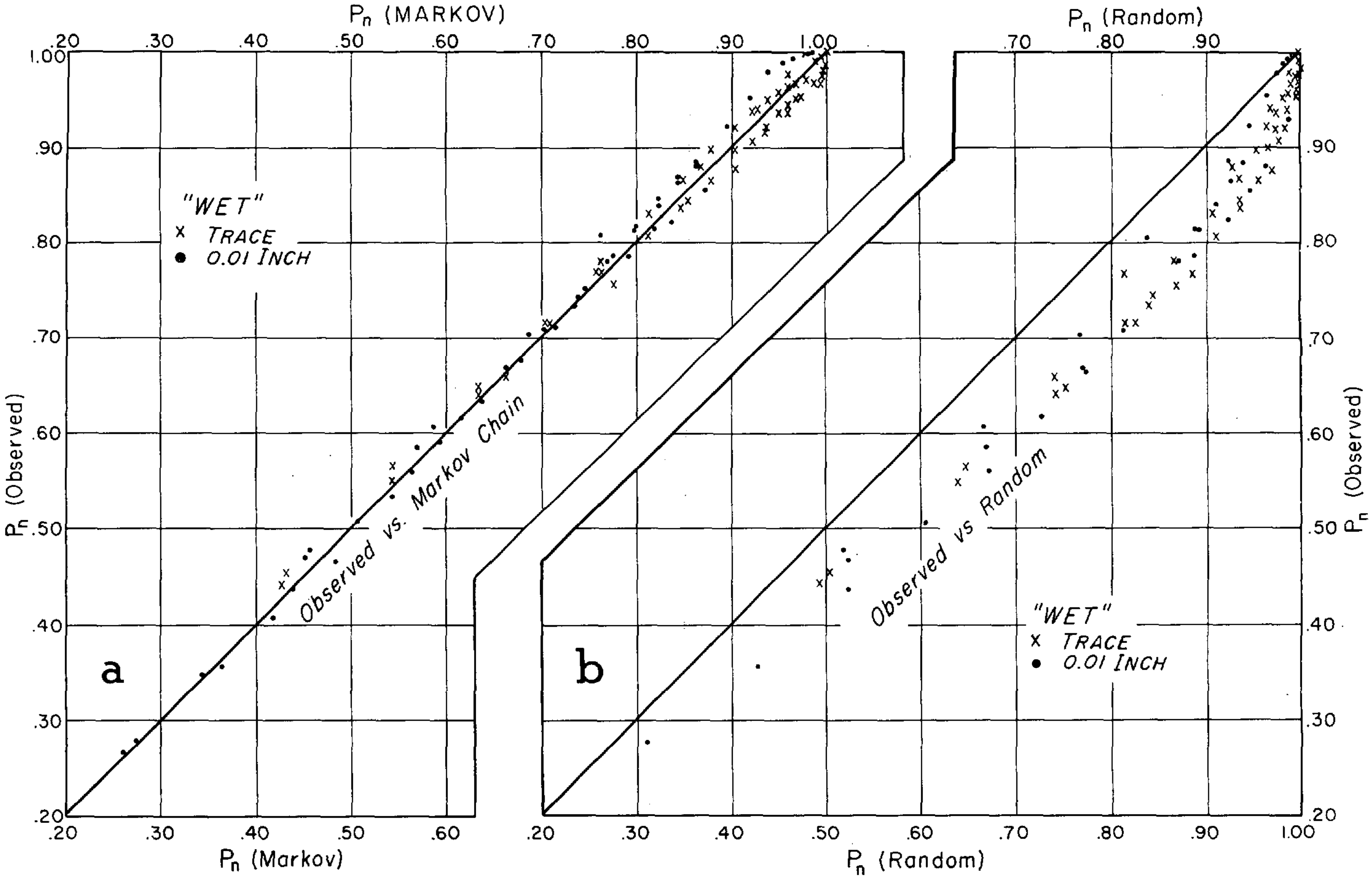

FIGURE 3.—Comparison of Topil's [1] observed relative frequency of precipitation occurrence with probability given by (a) Markov chain model and (b) random model. Values for the two definitions of precipitation occurrence ("wet") and for all integer values of length of period $n>1$ are plotted. Values for $n=1$ are excluded since P_1 is a parameter of the models.

note are needed to answer these and other questions on limitations of the utility of the simple Markov chain model for describing properties of precipitation statistics.

ACKNOWLEDGMENTS

The author is grateful to A. G. Topil for the opportunity to use his data while still in manuscript form and for supplementary data required to complete this study.

REFERENCES

1. A. G. Topil, "Precipitation Probability at Denver as Related to Length of Period," *Monthly Weather Review*, vol. 91, No. 6, June 1963, pp. 293–297.
2. K. R. Gabriel and J. Neumann, "A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv," *Quarterly Journal of the Royal Meteorological Society*, vol. 88, No. 375, Jan. 1962, pp. 90–95.
3. A. G. Topil, personal communication, July 31, 1962.
4. D. L. Jorgensen, "Persistency of Rain and No-Rain Periods During the Winter at San Francisco," *Monthly Weather Review*, vol. 77, No. 11, Nov. 1949, pp. 303–307.
5. A. H. Eichmeier and W. D. Baten, "Rainfall Probabilities During the Crop Season in Southern Lower Michigan," *Monthly Weather Review*, vol. 90, No. 7, July 1962, pp. 277–281.